Lecture 3

Physical layer: Data communications

References:

- CN chapter 2.

Purpose of physical layer

- For two parties to successfully communicate, the first thing needed is a **media** for communication to take place.
- The two parties must agree on a **common media**, and **what physical phenomenon in this media** is used to carry signal.
- It is also desirable to have a common standard for the computers to connect to the media, i.e., same shape, size, etc., of the connectors.
- They also need to agree on **a common form of signals**, e.g., what voltage, frequency, etc., to use.
- Then there must be mechanisms to **generate** and **detect** signals in the common media.

The physical layer cater for these basic needs. They are mostly of interest to electrical engineers than to computer scientists.

Being computer scientists, we will ignore the most "electrical" parts.

Copper wires as media

- **A pair of wires** can serve for a **point-to-point** connection: the sender connects a "voltage source" to it, the other end detects it.
- It can also be a good **broadcast** media: E.g.:
  - A voltage source provides some voltage between the wires.
  - The resistance between the wire is made high ("high impedance").
  - So normally everybody detects a high voltage. If one of them want to send, it shorts the wires, so everybody detects a 0 voltage.
  - We only have two signals: high or low. Computers contend for the media (chaos ensue if many computers want to send at the same time).

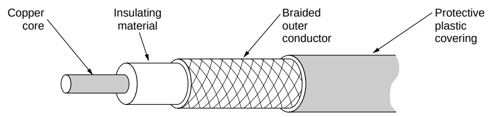
    How to prevent chaos? We will see it when we talk about "multiple access".
- Two wires running straight is a good antenna, meaning a **large loss of energy** and poor range. We need some way to prevent this.

UTP vs. Coaxial cable

- The wires can be **twisted.** This way **alternate** placement of wires generates **opposite magnetic fields**, cancelling each other.

- **Finer twisting** of these "Unshielded Twisted Pair" (UTP) gives **better performance** (less attenuation on higher frequencies).

  The above shows relative tightness of category 3 and category 5 UTP twisting. They are good for up to 16MHz and 100MHz respectively.
- An alternative: one wire is **completely within the other**. This way little magnetic field "leaks".



The catch: they are thicker and thus less flexible physically.

Visible light laser

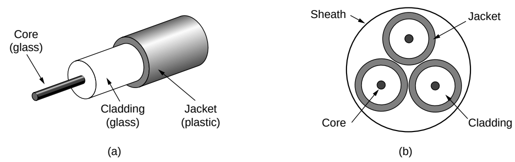
- Light is (nearly) unique in that it needs no media to travel.
- Again there are two states: has light and no light, usually has light represents a one-bit, no light represents a zero-bit.
- **Point-to-point** communication is used most of the time, although sometimes it is possible to implement broadcasting.
- Without a "guide media", light beam **lose energy** as the **square of distance**, and becomes **indistinguishable** from high **background noise** in moderate distance.

  That everything under the sun emits light is not helpful to communication.
- One way to deal with it: use **laser**. It has very little spread (i.e., highly uni-directional), so it loses energy much less seriously.
- Disadvantage: since it doesn't spread, it **must be arranged accurately**, and atmospheric disturbance can easily cause loss of connection.

Fiber cable

- A better solution: use something like a wire to "contain" the light.
- Fiber: glass being so **pure** that it is nearly **completely transparent**, so it works as a guiding media for long distance communication.
- **Total internal reflection** may be used to prevent energy to lost to the surroundings. I.e., angle of incident is large enough.

Side view of a fiber and cross-section view of a bundle of fiber:

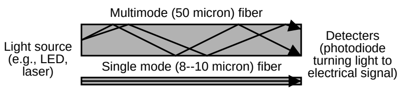


Multi-mode vs. single-mode

There is a problem: there are infinitely **many different angles** of incident that can result in total internal reflection, called different modes.

Paths of different modes has different lengths, so they transfer at different speed and interfere each other. One cure is to lower the transmission rate or length of the fiber .

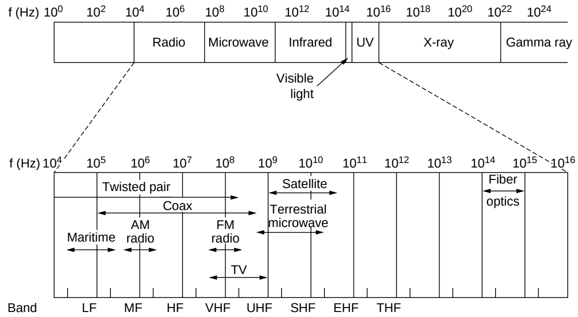
More expensive variant: **extremely fine** fiber that traps light within a few wavelength. The result is that light always follows exactly the path of the fiber ,although the signal is much weaker.



For single-mode fiber ,there is only one path. There is still a little bit of spreading (chromatic dispersion), but can be counteracted. It achieves 50Gbps signal rate over 100km without amplification (unluckily CPUs can't process data that fast).

Electromagnetic spectrum

Indeed, not just visible light can be used for communication. Any part of E-M wave can also do so, sometimes more conveniently.



Radio and Microwave transmission

- The first benefit is that there is lower noise than light, so it is possible to do without guiding media—**Wireless**.
- Second benefit: more **controllable frequency**. One can easily make a device that generates and detects radio wave with variable frequency.
  If you have controllable frequency electric signal, you just need an antenna to bring them into the atmosphere, and another to get it back to electrical signal.
- Limitation 1: To be easily generated and detected, one has to use a **narrow range of frequencies**. I.e., modulation must be used.
  Also, E-M wave of different freq's travel at slightly different speed in the air.
- Limitation 2: Large objects **obstruct EM wave propagation**. What is large depends on the wavelength (i.e., inverse of frequency).
  But this can be turned to opportunity: you can have more than one channel of the same frequency if every channel has its own direction.
- Limitation 3: it is **easy to tap**, and no way to detect tapping.
  We have to implement encryption somewhere.

Distance limitation

- **Attenuation**: Unguarded waves lose energy as square of distance. Guarded waves and voltage lose energy to the media since it's not completely transparent (although the loss is "only" proportional to distance).
- **Distortion due to speed variation**: at different frequencies, wave and voltage travel at different speed, and signals at close times will interfere each other. This limits the signal rate.
  Square waves and aperiodic signals should be considered to be the sum of many sine waves, so they are affected as well, for fix edfrequency pulses.
- **Distortion due to attenuation variation**: at different frequencies, wave and voltage attenuate at different rate, so it becomes more difficult to recognize.
- **Noise**: the amount of random distortion is proportional to the length of transmission.
- **Delay**: it takes a long time to know about transmission error.

Concept of repeaters

Solution: add intermediate receivers and transmitters.

- **Stationary**: radio station, Geostationary Earth Orbit (GEO) satellites, electrical "repeaters", optical regenerators, etc. Simpler to operate.

  Most communication satellites are GEOs, and the orbit should be considered full. Fibers are generally connected by terrestrial regenerators.
- **Moving**: Medium and low Earth Orbit (MEO, LEO) satellites. More difficult to operate. The GPS system operates using MEOs, and a project called Iridium uses LEOs.
  Difficult because one must find another satellite when one fall out of range. Many of them must work together to provide full coverage: 24 for MEO, 70 for LEO.
- **Hybrid**: A moving satellite serves only to relay the message to a ground station for terrestrial repeating. Useful for **mobile** communication. Used by the Teledisc project.

Trends

- **Fiber** has far the highest **bandwidth**, but it must be guided. It can be used only if the communicating parties are stationary.
- **Radio** has no such restriction, but the bandwidth is somewhat lower. They are thus mostly used when at least one of the communicating parties are **mobile**.
- **UTPs** has similar low bandwidth as radio, but they are **easily available** since most buildings have a lot of extra telephone wires which can be used. They are mostly used for connecting computers at a home to a fiber network.
- Small **LANs** also use **UTPs** and (less commonly) **coaxial cables**, but fibers are replacing them, especially for high volume traffic.

So in short, everything is moving towards wireless and fiber ,although UTP is still a prominent technology.

Representing signals

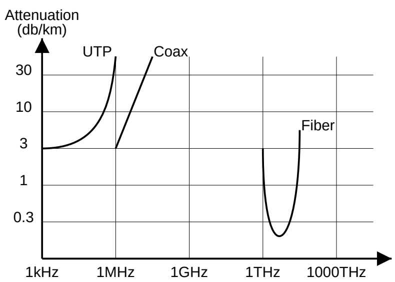
- When performing digital communication, the sender has **a sequence of bits to transfer** to the receiver.
  Digital is preferable to analog, as noise can be removed rather than accumulate.
- We already mentioned that we can **use voltages to represent bits**: a 0-bit might be represented by low voltage, 1-bit by high voltage.
- A sequence of bits can be represented by a sequence of voltages, **changing over time**.
  The sender and receiver must agree on some timing constraints, e.g., how long each voltage will last for.
- We call this **base-band** communication. It is characterized by the use of **channel frequencies** that **match the signal rate**.

Frequency and time domain

- As we know, every **time-varying** signal also has a **frequency domain** represention: they can be treated as **the sum of sine waves** of different frequencies, amplitudes and phases.
  Where frequencies can also be interpreted as changing phase.
- We've seen that attenuation and transmission speed depends on **frequencies** (not voltage), and thus has a tendency to distort the signal.
  If attenuation is too large, signal won't pass through. E.g., phone lines block dc.
- **Noise** is also frequency dependent.
- If we want longer distance (i.e., less repeaters), we must use **narrow band of frequencies**, which must be good for that media.
  But… our signal is probably not of that frequency!
- The size of range of frequencies used is called **bandwidth**. E.g., a channel using 2.410GHz to 2.415GHz has a bandwidth of 5 MHz.
  This is the first channel of a modern wireless LAN card.

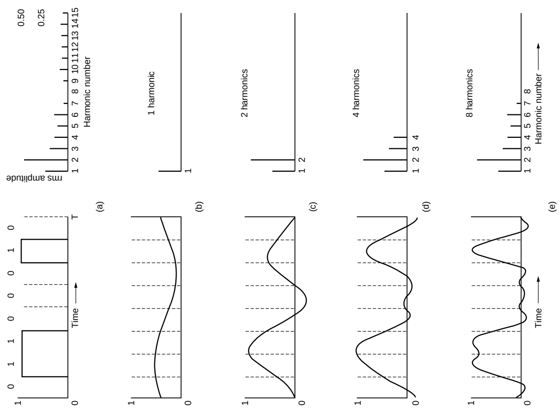
Channel characteristics

To see how large is the problem, let's see how much attenuation we can expect in common guided media.



From "Data and Computer Communications", William Stallings. An (imprecise) approximation, considering that there are many types of UTPs and many type of Coax.

Effect of limited spectrum



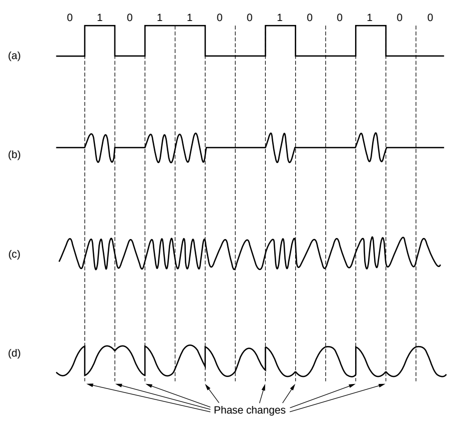
Here first harmonic has wavelength being $1/T$, not $1/8T$. E.g., if we use 4 harmonics, then we get 2bps per Hz.

Modulation ("Keying")

Some media allows **modulation,** matching different frequencies between the signal to transmit and the usable channel.

- A **carrier**, i.e., a single frequency wave for transmission, is generated.
- The characteristic of the carrier is **modified slightly** at a constant rate ("baud rate"). Each modification represent a signal.
  The modification should be recognizable using only a narrow spectrum.
- A number of characteristics of the sine wave can be modified:
  - **Amplitude** Shift Keying (ASK). Turning on and off the light source for fiber can also be interpreted as ASK
  - **Frequency** Shift Keying (FSK).
  - **Phase** Shift Keying (PSK), i.e., introducing "jump".
    "Keying" emphasize that the modulation wave is discrete rather than continuous.

Illustration



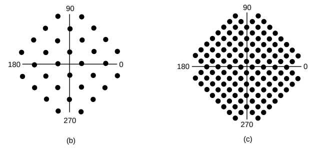
(a) baseband, (b) ASK, (c) FSK, (d) PSK

Why so many choices

- The **choice of modulation method is very complicated**, and is not in scope of our course.
- Here are some of the complications:
  - We want **ease of generation and detection**, so that devices can be cheaper. This favours ASK.
  - We also want **low power consumption**. Again, favours ASK.
  - We want **low bandwidth usage**. Again, favours ASK.
  - We want **high noise immunity**. This favours FM and PSK.
  - We want **ability to represent more than 2 different values** in the same baud. This favours ASK and PSK.
- And the list goes on. But it is impossible to fully understand them without deep mathematical background, so we will skip them.

Multiple characteristics

- Multiple characteristics can be used simultaneously to represent **more symbols at the same baud**.
  Rationale: by using multiple characteristics, the number of different values for each is less, so there is more difference between them, improving reliability.
- E.g., modems use a **combination of AM and PSK**. E.g., 9.6kbps and 14.4kbps modems use 2400 baud, so they send 4 and 6 data bits per baud. Adding 1 bit for error detection, they need 32 and 128 symbols.
  Phone lines cutoff at 3kHz, so 2400 baud is about the most they can do.



"Constellation diagram": each dot represent a symbol. The distance from center is the amplitude, the angle from 0 shows the phase.

Theoretical limits

- Can we **infinitely add more and more dots** into the constellation diagram to make transmission faster and faster, using the same bandwidth? Answer: Without noise, yes. With noise, no.
- **Shannon** (1948) proving the following sharp relation between noise and bit rate achievable using any modulation (Shannon bound):

For a channel of H Hz, where signal strength and noise strength is in the ratio S/N, we can transfer no more than $H\log_2(1+S/N)$ bits in each second, regardless of modulation methods.

  The proof assumes thermal noise, i.e., noise follows Gaussian distribution. Other common distributions doesn't change the results significantly ,however. It requires only information theory, so violation is impossible.
- So for phone systems, $H = 3000$, S/N is around 1000, so we can send only around 30000 bits per second.
  Normally noise is counted in "decibels (dB)", each 10dB (or bel) meaning 10 more times. So 1000 times is 3 bels, i.e., 30dB.

Frequency characteristic of modulations

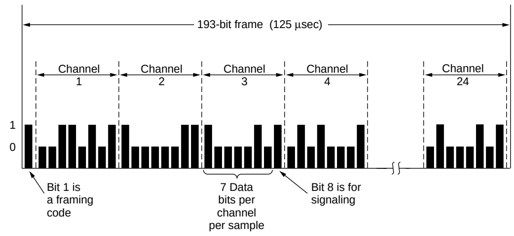
Let's get back to our bottom line for a short while.

- To use modulation, we generate a wave of frequency f as carrier.
- We use a signal of frequency f' to modify the carrier.
  In principle, the frequencies is infinite for discrete signals. But we will just use a few harmonics, the frequency should be less than the bit rate (see 3.15).
- **What frequencies we need** to represent the result?
- It is natural to expect frequency f is needed.
- What else? It **depends on the modulation**. E.g., ASK signal with k harmonics would use $f - f'$ to $f + f'$, occupying a bandwidth of $2f'$.
  It is possible to remove half of it, though.
- So the bandwidth needed depends on f', which in turn depend on the **baud rate** and **how many harmonics** to keep.
  Which then depends on the amount of noise in the channel.

Multiplexing, TDM

- There is **vast difference** of bandwidth limit on different media.
- E.g., a phone user talked into his phone, and at the toll office of the phone company the 3kHz bandwidth is digitized into 56kbps.
  56kbps is 128 levels (i.e., 7-bits) quantization sampled at 8kHz. Such analog-to-digital conversion is known as Pulse Code Modulation (PCM).
- But the connections between toll offices **far exceed** 56kbps!
- Natural idea: use the connection to serve **many** users, each of them talk **at different time**: **Time** domain multiplexing (TDM).
- E.g., T1 (1.544Mbps) scheme:
  - The channel allows transfering 8000 groups of 193 bits each second.
  - Each user needs only 56kbps, so needing 7 bits only. Adding a control bit, we can serve 24 simultaneous users (and has 1 bit left for signalling and synchronization).

T1 signals



Signalling bits are used for connection, disconnection, routing, synchronization, etc. Perhaps 8kbps per channel is too much, though.

This signal is directly sent to the other end of the twisted pair, with repeaters inserted at regular interval to amplify signal and remove noise.

Why even at T1 (24 voice channels) we can do base-band signaling, while in modem we have to invent complicated modulation scheme? Answer: our phone line has bandwidth artificially limited to voice band.

Limitations

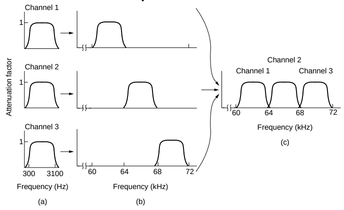
PCM+TDM is simple, but is next to impossible to use on occasions.

- Broadcast media: different receivers receive at different time (due to difference distance), so it's **hard to synchronize**.
  E.g., wireless radio, Ethernet. It is better if everyone is talking to the same central station, e.g., GSM in mobile phone uses some TDM.
- For very fast communications, **computers are too slow** to combine the bits coming from the channels.
  E.g., optic fiber . The full bandwidth of a fiber is around 100THz (See 3.14), where is the computer that can switch at that rate?

One alternative: simply send at different frequency bands of the channel.

FDM and WDM

- Different frequencies of the spectrum **won't interfere** each other:



- They can be added together and later **separated** from each other easily using **band-pass filter** (electrical) or **interferometer** (optical).
- We call this **wavelength** (WDM) or **frequency domain multiplexing** (FDM), depending on whether we are talking about optical or not.
  But they are exactly the same thing: don't forget the lighting and not lighting is actually just ASK.

Noise immunity and spread spectrum

- Noisy is usually very **time and frequency specific**. This is no good news for both TDM and FDM.
- For TDM, a burst of noise will **corrupt a few bits of all channels** completely, all will start waiting for timeout and retransmission.
- For FDM, noise in a specific frequency **render a few channels useless completely**, even though other channels can still work.
- Solution: communicate in **multiple** channels in increase noise immunity: **spread spectrum**.
- This **increases** bandwidth consumption, which is a really bad news. Can multiple users shares a wide spectrum and thus share the noise prevention cost?
- Answer: **Code Division Multiple Access** (CDMA).
  Employed in radio communication like mobile phones and network links.

CDMA sending

- Requirement: **additive channel**. Radio provides a good one (still remember constructive and destructive interference?).
- Scheme: every channel is allocated a unique *m* bit **chip sequence**. Usually $m = 64$ or 128, but we will use $m = 4$ to simplify discussions.
  A *m* bit chip allows at most *m* simultaneous users. *m* must be a power of 2.
- Each chip sequence is interpreted as a sequence of -1 and 1. For our example: [1, 1, 1, 1], [1, -1, 1, -1], [1, 1, -1, -1], [1, -1, -1, 1].
- Time is **slotted**. During each slot, each station can choose one of 3 things to do at each slot: send nothing, 1 or -1.
- To send 1, the chip message is sent. To send -1, the negation of that is sent. To send nothing, *m* bits of 0's are sent.
  This is described in base-band. In radio, we cannot send in base-band, so one instead send a sine wave or a negated sine wave, i.e., do PSK, depend on the number within the chip.

Example

E.g., suppose three of the channels want to send a bit in a slot. The first want to send 1, the second 1, and the fourth -1. What will happen?

- Message sent by 1st station: [1, 1, 1, 1].
- Message sent by 2nd station: [1, -1, 1, -1].
- Message sent by 4th station: [-1, 1, 1, -1].
- Channel is additive, so this results in [1, 1, 3, -1].

How the receiver **seperate** them? The technique is **corelation** you know in statistic: test whether the result corelate with the chip sequence.

E.g., for 2nd channel: correlation bewteen [1, -1, 1, -1] and [1, 1, 3, -1] is $1 \times 1 + -1 \times 1 + 1 \times 3 + -1 \times -1 = 4$.

Being very positive, we conclude that a 1 was sent in channel 2.
The result should be one of m, $-m$ and 0.

Why it works?

- There is a special property of the 4 chip sequences we choose: **they have zero corelation**, i.e., "**orthogonal**".
  That's why we only have *m* channels rather than 2^m: we got only that many orthogonal sequences.
- So channel 2 contributes 4 to the final corelation, while all other channels contribute nothing. The sum is of course 4.
  Actually we can send things other than just 1 and -1. Again, the limitating factor is noise.
- If some noise makes one bit from 1 to 0, the sum may decrease from 4 to 3, and an idle channel might change from 0 to 1. Still distinguishable.
- So if we have more bits in a chip (e.g., 64), only a very large noise or one which last for very long could corrupt our channel.
- **In terms of bandwidth**, we increase bandwidth 64 times and let 64 channels share that bandwidth, to increase noise immunity 64 times.